DOCUMENT RESUME

ED 126 113 TM 005 368

AUTHOR Tierney, Robert J.; And Others
TITLE The Differential Diagnostic Properties of the Revised Illinois Test of Psycholinguistic Abilities and the Wechsler Intelligence Test for Children-Revised.
PUB DATE [Apr 76]
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Children; Comparative Analysis; Comparative Testing; Cross Cultural Studies; *Diagnostic Tests; *Educational Diagnosis; Elementary Secondary Education; Evaluation; Individual Differences; *Intelligence Tests; Language Tests; *Psycholinguistics; Standardized Tests; *Statistical Analysis; Test Validity
IDENTIFIERS Australia; *Illinois Test Psycholinguistic Abilities Revised; United States; *Wechsler Intelligence Scale for Children Revised

ABSTRACT

The topic of differential diagnosis and its application to the use of the Revised Illinois Test of Psycholinguistic Abilities (ITPA) and Wechsler Intelligence Scale for Children-Revised (WISC-R) is discussed. The theoretical bases for differential diagnosis and the properties of assessment instruments with diagnostic properties are presented. Statistical procedures for determining whether tests have differential diagnostic power are discussed and illustrated by using the American and Australian standardization data from the ITPA and the American standardization data from the WISC-R. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED126113

BEST COPY AVAILABLE

TM005-368

The Differential Diagnostic Properties of the Revised Illinois Test of Psycholinguistic Abilities and the Wechsler Intelligence Test for Children - Revised.

Authors: Robert J. Tierney, University of Arizona
Wilbur S. Ames, University of Arizona
Robert Teasdale, The Flinders University of South Australia

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

The differential diagnosis of the ability of an individual pertains to the discovery of their strengths and weaknesses. For purposes of education, its worth is self evident. Differential diagnosis provides information which acts as the basis for guiding the educational experiences of an individual. Ideally, it makes possible a more accurate program of instruction. Realistically, test constructors rarely provide adequate evidence as to the efficiency of their tests as instruments of differential diagnosis. Therefore, implementation of a diagnostic prescriptive ethic via differential diagnosis must remain at best, presumptuous. In other words, as few test constructors provide evidence of the differential diagnostic efficiency, decisions relative to both the general and specific directions of educational endeavors must remain either inaccurate or, purely speculative.

Given this predicament, the present paper has a two-fold purpose. First, it is concerned with a delineation of methods for determination of the differential diagnostic efficiency of tests. Second, it is concerned with a description of the results derived from the application of such procedures to the assessment of the differential diagnos-

tic efficiency of two instruments which are widely used throughout the world for diagnostic purposes. Specifically, the differential diagnostic efficiency of the Revised Illinois Test of Psycholinguistic Abilities (ITPA) and Wechsler Intelligence Scale for Children-Revised (WISC-R) are evaluated.

Determining the Efficiency of Tests to Make Differential Diagnosis

Determination of the differential diagnostic efficiency of tests is synonymous with an evaluation of the validity with which tests can be used as agents of differentiation. More specifically, this aspect of validity of a test is concerned with whether the differences between test scores actually entails a real difference between the skills and/or abilities that tests purport to measure. Assuming the skills and/or abilities are represented in the various tests within a battery, the basic question to be answered is: Does the test battery exhibit the differential diagnostic efficiency necessary for valid subtest comparisons?

There are various indices by which a test constructor and test user can assess the differential diagnostic value of an instrument. For example, reliability of difference procedures (Mosier, 1951), an assessment of attentuation (Nunnally, 1967), or Kelley's method (Kelley, 1923) for estimating the proportion of differences in excess of chance might all be considered appropriate. However, in terms of defining whether a comparison between scores on two tests is legitimate, Kelley's method for determining the proportion of differences in excess of chance has the advantage of being directly meaningful. Based on the standard deviation of differences due to the unreliability of the

measures and the standard deviation of the obtained distribution of differences, Kelley's index provides a determination of the proportion of differences between scores on a pair of tests which can not be accounted for by chance. For example, for two tests a proportion of differences in excess of chance of .32 indicates that, regardless of the size of the differences, only 32 per cent of all the differences observed between the two tests would represent true differences. This would be due both to the imperfect reliabilities of the two tests and to the intercorrelation between the tests.

In terms of evaluating the diagnostic value of test comparisons, for which the proportion of differences in excess of chance had been assessed, the present investigators were concerned with a determination of what proportion of chance differences should be considered minimal. A review of past studies revealed that a proportion of differences in excess of 25 per cent was considered minimal. In 1934, Segel in his research with the Kelley procedure, originated this criterion. He offered the judgement that any two tests exhibiting a proportion of differences in excess of chance less than this value should not be considered diagnostic in relation to each other or, in other words, should not be used in differential diagnosis. The arbitrariness with which Segel derived this criterion must be acknowledged; however, it must be recognized that requiring at least one-fourth of the differences observed between two tests to be true differences does not seem an unreasonable prerequisite. Indeed, it would seem a minimum. In terms of evaluating the diagnostic value of test comparisons, the present investigators considered Segel's 25 per cent crucial value

an appropriate criterion for use in their analyses. Bennett and Doppelt (1948) accepted and used this same value in their analyses of tests.

Differential Diagnosis and the Revised Illinois Test of Psycholinguistic Abilities

The ITPA has been used extensively throughout the world for the purpose of assessing the strengths and weaknesses of children who are experiencing learning disabilities. However, if we are to legitimately use the ITPA to pinpoint intra-individual differences, an examination and appreciation of the differential diagnostic efficiency of the ITPA is essential.

Given the nature and function of the ITPA, the question to be answered is: Does the ITPA exhibit the differential diagnostic efficiency required for subtest comparisons at various age levels? For this purpose, the Kelley procedure for estimating the proportion of differences in excess of chance was applied to the United States' standardization data. Then, Segel's criterion was used to evaluate the viability of each subtest comparison at the various age levels. To be consistent with the data used in the development of the ITPA norms and derivation of the standard error of measurement of differences, the standardization data, which were uncorrected for restricted intelligence range, were used in the present analyses.

Table 1 contains information concerning which subtest comparisons at each age level were found to have insufficient differential diagnostic efficiency. A tabulation of the data in table 1 reveals the extent of the problem. Forty-five of the possible 66 subtest compari-

sons were isolated as being inappropriate at least at one age level. Two of the comparisons (Auditory Association-Grammatic Closure; Grammatic Closure-Auditory Closure) were found to be inappropriate at all age levels; a total of 14 subtest comparisons were found to be inappropriate at more than half their age levels. At one age level, as few as 8 comparisons were appropriate.

Insofar as evaluating the differential diagnostic potential of the ITPA, the preceding data suggest that considerable caution should be exercised by anyone wishing to use the ITPA for diagnostic purposes. Minimally, reference to the data in table 1 should precede any attempt to use the ITPA for purposes of differential diagnosis. If, according to these data, certain subtest comparisons are tenable, then appropriate confidence bands can be applied to individual score data. For example, Paraskevopoulos and Kirk (1969) suggest the use of 2.5 SE_m confidence bands. Alternatively, if the data in table 1 suggest that certain subtest comparisons are not tenable, then attempts at differential diagnosis using these comparisons should be rejected.

Given the widespread use of the ITPA throughout much of the world a decision was made to extend these analyses to include an examination of the differential diagnostic use of the ITPA with a population of different national origin. To this end, comparable, individual raw score data for an Australian sample at two age levels were derived (Teasdale and Wray, 1974). This analysis proceeded in two stages. First, in order to assess the suitability of selected ITPA subtest items across cultures, a systematic examination, via means of direct comparisons, of the Australian samples performance on individual items

of selected ITPA subtests was undertaken. Thereupon, if proven appropriate, the second stage would entail an analysis of the differential diagnostic efficiency of ITPA subtests for Australian subjects, based upon these Australian standardization data.

The first stage of this analysis involved an assessment of the difficulty level of the items for seven of the ITPA subtests. Five ITPA subtests were excluded either because they do not lend themselves to conventional item analysis procedures (Manual Expression, Verbal Expression and Visual Closure) or, because they have somewhat of a "built-in" item-test difficulty (Auditory Sequential Memory and Visual Sequential Memory).

The results of these item analyses are displayed graphically by age levels and countries in figures 1 through 14. Overall, the distributions of item-test difficulties reveal several vacillations. Without much consistency, item difficulties vary as a function of country and age level. In selected subtests, the distribution of item test difficulties for specific age levels and/or countries are characterized by several extreme fluctuations in difficulty level.

The erratic nature of these distributions raise several questions. Variations across country and age levels question the generalizability of selected ITPA subtest items as measures of psycholinguistic abilities either across countries or, across age levels within countries. The fluctuations in difficulty level from item to item question the appropriateness with which items within selected subtests have been ordered. (A more detailed discussion of these results and their implications is available in a recent paper by Teasdale, Tierney, Wray and Ames.)

Given the results of this preceding analysis, the present investigators deemed it inappropriate to proceed with an analysis of the differential diagnostic potential of the ITPA subtests for Australian subjects. Because of the obvious need for major revisions of the ITPA, any attempt to further examine and detail the differential diagnostic properties of the ITPA might support its continued use in Australia without such modifications.

Differential Diagnosis and the Wechsler Intelligence Scale for Children-Revised

As with the ITPA, the Wechsler Intelligence Scale for Children has been an extremely popular instrument for examining intra-individual differences via comparisons across subtests. Ferinden and Jacobson (1969) and Glasser and Zimmerman (1967) detail procedures for interpreting subtest performances. More recently, Searles (1975) suggested that the subtest scores were the most important scores the WISC yield. As it would seem valid to assume that the recently revised WISC will receive similar attention, an examination of its differential diagnostic properties would seem integral to assessing the viability with which such diagnoses can proceed.

The basic question to be answered relative to the Wechsler Intelligence Scale for Children-Revised (WISC-R) is: Does the WISC-R exhibit the differential diagnostic efficiency necessary for subtest comparisons at various age levels? For this purpose, Kelley's procedure for estimating the proportion of differences in excess of chance was applied to the North American standardization data for each subtest and Siegel's criterion used to evaluate the viability of the comparisons.

Table 2 contains the information concerning which subtest comparisons at each age level were found to have insufficient efficiency for purposes of differential diagnosis. Tabulation of these data reveal that only 47 per cent of the possible comparisons across subtests at the various age levels are of some diagnostic value. For selected tests, less than 10 per cent of their possible comparisons are tenable.

Given the limitations of the differential diagnostic properties of the WISC-R, its extensive use for purposes of differential diagnosis is questionable. Unless persons using the test exert considerable caution, their attempts at pinpointing an individual's strengths and weaknesses will likely be misguided. Minimally, any attempt at differential diagnosis would require a person's referencing and considering the information presented in table 2.

The widespread use of the WISC throughout the world underlines the need for an analysis of its use diagnostically with persons of various national origins. Unfortunately, the lack of available data prevented the present investigators from undertaking such an assessment.

Summary and Conclusions

With the increased emphasis upon a diagnostic prescriptive ethic and the subsequent widespread use of diagnostic tests, it is hoped that both test constructors and test users will be more sensitive to the diagnostic properties of their measures. To this end, the present paper has served three purposes. First, a case for assessing the differential diagnostic power of so-called "diagnostic" tests has been presented. Second, the application of fundamental procedures for

assessing the efficiency of selected diagnostic batteries has been demonstrated. Third, the findings and implications relative to the differential diagnostic efficiency with which the ITPA can be used with both Australian and North American populations, and with which the WISC-R can be used with North American populations, has been detailed. The case presented, the procedures suggested and the findings of these analyses have ramifications for constructors and users of tests - especially of the ITPA and WISC-R.

REFERENCES

Bennett, George K. and Doppelt, Jerome E. "The Evaluation of Pairs of Tests for Guidance Use." Educational and Psychological Measurement, 1948, 8, 319-325.

Ferinden, William E., Jr., and Jacobson, Sherman. Educational Interpretations of the Wechsler Intelligence Scale for Children (WISC). Linden, New Jersey: Remediation Associates, 1969.

Glasser, A. J., and Zimmerman, I. L. Clinical Interpretation of the Wechsler Intelligence Scale for Children. New York: Grune and Stratton, 1967.

Kelley, Truman L. "A New Method for Determining the Significance of in Intelligence and Achievement Test Scores." Journal of Educational Psychology, 1923, 14, 321-333.

Kirk, S. A., J. J. McCarthy and W. D. Kirk. Revised Illinois Test of Psycholinguistic Abilities. Urbana, Illinois: University of Illinois Press, 1968.

Mosier, C. J. Batteries and Profiles In E. F. Lindquist (Ed.) Educational Measurement. Washington, D. C.: American Council on Education, 1951, 764-808.

Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.

Paraskevopoulos, J. N. and S. A. Kirk. The Development and Psychometric Characteristics of the Revised Illinois Test of Psycholinguistic Abilities. Urbana, Illinois: University of Illinois Press, 1969.

Searls, Evelyn F. How To Use WISC Scores In Reading Diagnosis. Newark, Delaware: International Reading Association, 1975.

Segel, David. Differential Diagnosis of Ability in School Children. Baltimore: Warwick and York, 1934.

Wechsler, David. Manual for the Wechsler Intelligence Scale for Children - Revised. New York: The Psychological Corporation, 1974.

Table 1 ITPA Subtest Comparisons with Proportion of Differences in Excess of Chance Less Than 25%

Subtest - Pair	Age Level 2.7 to 3.1	3.7 to 4.1	4.7 to 5.1	5.7 to 6.1	6.7 to 7.1	7.7 to 8.1	8.7 to 9.1	9.7 to 10.1
Auditory Reception - Verbal Expression	*							
Auditory Reception - Grammatic Closure	*	*	*				*	*
Auditory Reception - Visual Closure	*	*						
Auditory Reception - Auditory Closure							*	*
Verbal Reception - Auditory Association						*		*
Visual Reception - Verbal Expression	*			*				
Visual Reception - Grammatic Closure	*	*		*			*	*
Visual Reception - Visual Closure	*	*	*				*	*
Visual Reception - Visual Sequential Memory					*		*	
Visual Reception - Auditory Closure							*	*
Auditory Association - Verbal Expression	*			*				
Auditory Association - Manual Expression								*
Auditory Association - Grammatic Closure	*	*	*	*	*	*	*	*
Auditory Association - Visual Closure	*	*	*	*		*		*
Auditory Association - Visual Sequential Memory								*

Table 1 (continued)

Subtest - Pair	Age Level 2.7 to 3.1	3.7 to 4.1	4.7 to 5.1	5.7 to 6.1	6.7 to 7.1	7.7 to 8.1	8.7 to 9.1	9.7 to 10.1
Auditory Association - Auditory Sequential Memory	*							
Auditory Association - Auditory Closure				*	*		*	*
Visual Association - Verbal Expression	*			*				
Visual Association - Grammatic Closure		*		*		*	*	*
Visual Association - Visual Closure	*	*	*	*	*			
Visual Association - Visual Sequential Memory								*
Visual Association - Auditory Closure							*	*
Verbal Expression - Manual Expression	*		*	*	*		*	
Verbal Expression - Grammatic Closure	*	*	*	*			*	
Verbal Expression - Visual Closure	*	*	*	*	*		*	
Verbal Expression - Auditory Sequential Memory	*	*						
Verbal Expression - Auditory Closure	*			*			*	*
Verbal Expression - Sound Blending	*	*						
Manual Expression - Grammatic Closure	*	*						
Manual Expression - Visual Closure	*	*					*	

Table 1 (continued)

Subtest - Pair		Age Level 2.7 to 3.1	3.7 to 4.1	4.7 to 5.1	5.7 to 6.1	6.7 to 7.1	7.7 to 8.1	8.7 to 9.1	9.7 to 10.1
Manual Expression	- Auditory Closure							*	*
Grammatic Closure	- Visual Closure	*	*	*	*		*	*	*
Grammatic Closure	- Visual Sequential Memory		*		*	*			*
Grammatic Closure	- Auditory Sequential Memory	*	*						
Grammatic Closure	- Auditory Closure	*	*	*	*	*	*	*	*
Grammatic Closure	- Sound Blending	*	*						
Visual Closure	- Auditory Sequential Memory	*	*						
Visual Closure	- Visual Sequential Memory		*	*	*	*			*
Visual Closure	- Auditory Closure	*	*	*		*	*	*	*
Visual Closure	- Sound Blending	*	*	*					
Auditory Sequential Memory	- Visual Sequential Memory					*			
Visual Sequential Memory	- Auditory Closure				*	*		*	*
Visual Sequential Memory	- Sound Blending					*			
Auditory Closure	- Sound Blending							*	

Table 2 WISC-R Subtest Comparisons With Proportion of Differences in Excess of Chance Less than 25%

Subtest Comparison	Age Level										
	6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Information - Similarities	*	*	*	*	*	*	*	*	*	*	
Information - Arithmetic	*		*	*	*	*			*	*	*
Information - Vocabulary	*	*	*	*	*	*	*	*	*	*	*
Information - Comprehension	*	*	*	*	*	*		*	*	*	*
Information - Digit Span	*			*							
Information - Picture Completion	*				*	*	*		*		
Information - Picture Arrangement	*	*	*	*				*			
Information - Block Design	*	*		*							
Information - Object Assembly	*	*	*	*	*		*				
Information - Coding	*										
Information - Mazes	*						*				*
Similarities - Arithmetic			*	*	*	*	*	*	*	*	*
Similarities - Vocabulary	*	*	*	*	*	*	*	*	*	*	*
Similarities - Comprehension	*	*	*	*	*	*	*	*	*	*	*
Similarities - Digit Span			*	*	*						
Similarities - Picture Completion			*	*	*	*	*	*	*	*	*

Table 2 - Continued

Subtest Comparison	Age Level 6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Similarities - Picture Arrangement		*	*	*	*			*	*	*	
Similarities - Block Design				*					*	*	
Similarities - Object Assembly		*	*	*	*	*	*		*	*	*
Similarities - Coding		*						*	*		
Similarities - Mazes				*	*		*			*	*
Arithmetic - Vocabulary	*	*		*	*				*		
Arithmetic - Comprehension	*	*	*	*	*			*	*	*	*
Arithmetic - Digit Span	*	*	*	*	*		*	*	*		*
Arithmetic - Picture Completion					*			*	*	*	
Arithmetic - Picture Arrangement	*	*	*						*		
Arithmetic - Block Design	*	*		*	*		*		*		
Arithmetic - Object Assembly		*	*		*		*		*		
Arithmetic - Coding	*	*	*					*	*		
Arithmetic - Mazes					*		*				*
Vocabulary - Comprehension	*	*	*	*	*	*		*	*	*	*
Vocabulary - Digit Span	*			*				*			

Table 2 - Continued

Subtest Comparison	Age Level 6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Vocabulary - Picture Completion	*	*			*		*				
Vocabulary - Picture Arrangement	*	*	*	*							*
Vocabulary - Block Design	*	*		*							
Vocabulary - Object Assembly	*	*		*	*	*	*				
Vocabulary - Coding	*	*						*			
Vocabulary - Mazes							*				
Comprehension - Digit Span	*			*							
Comprehension - Picture Completion		*		*	*		*		*	*	*
Comprehension - Picture Arrangement	*	*	*	*	*			*	*	*	*
Comprehension - Block Design	*	*		*						*	
Comprehension - Object Assembly	*	*	*	*	*	*	*		*	*	*
Comprehension - Coding	*	*						*			
Comprehension - Mazes	*	*	*		*					*	*
Digit Span - Picture Completion											
Digit Span - Picture Arrangement	*										
Digit Span - Block Design				*							

Table 2 - Continued

Subtest Comparison	Age Level 6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Digit Span - Object Assembly				*	*		*				
Digit Span - Coding	*			*	*			*			
Digit Span - Mazes					*		*			*	*
Picture Completion - Picture Arrangement	*			*	*	*	*	*	*	*	*
Picture Completion - Block Design				*	*		*	*	*	*	*
Picture Completion - Object Assembly		*	*	*	*	*	*	*	*	*	*
Picture Completion - Coding	*							*	*		
Picture Completion - Mazes				*	*		*	*	*	*	*
Picture Arrangement - Block Design	*	*	*	*				*	*	*	
Picture Arrangement - Object Assembly	*	*	*	*	*	*	*	*	*	*	*
Picture Arrangement - Coding	*	*	*					*	*		
Picture Arrangment - Mazes			*	*	*		*	*	*	*	*

Table 2 - Continued

Subtest Comparison	Age Level 6½	7½	8½	9½	10½	11½	12½	13½	14½	15½	16½
Block Design - Object Assembly	*	*	*	*	*	*	*	*	*	*	*
Block Design - Coding	*	*						*			
Block Design - Mazes		*		*	*		*	*		*	*
Object Assembly - Coding	*	*	*		*		*	*	*		
Object Assmebly - Mazes	*	*	*	*	*			*	*	*	*
Coding - Mazes		*						*	*		*

Tables 1 and 2:
Code: * denotes a subtest comparison with proportion of differences in excess of chance less than 25 per cent.

Figure 1
AUDITORY RECEPTION - 5.7 - 6.1 years

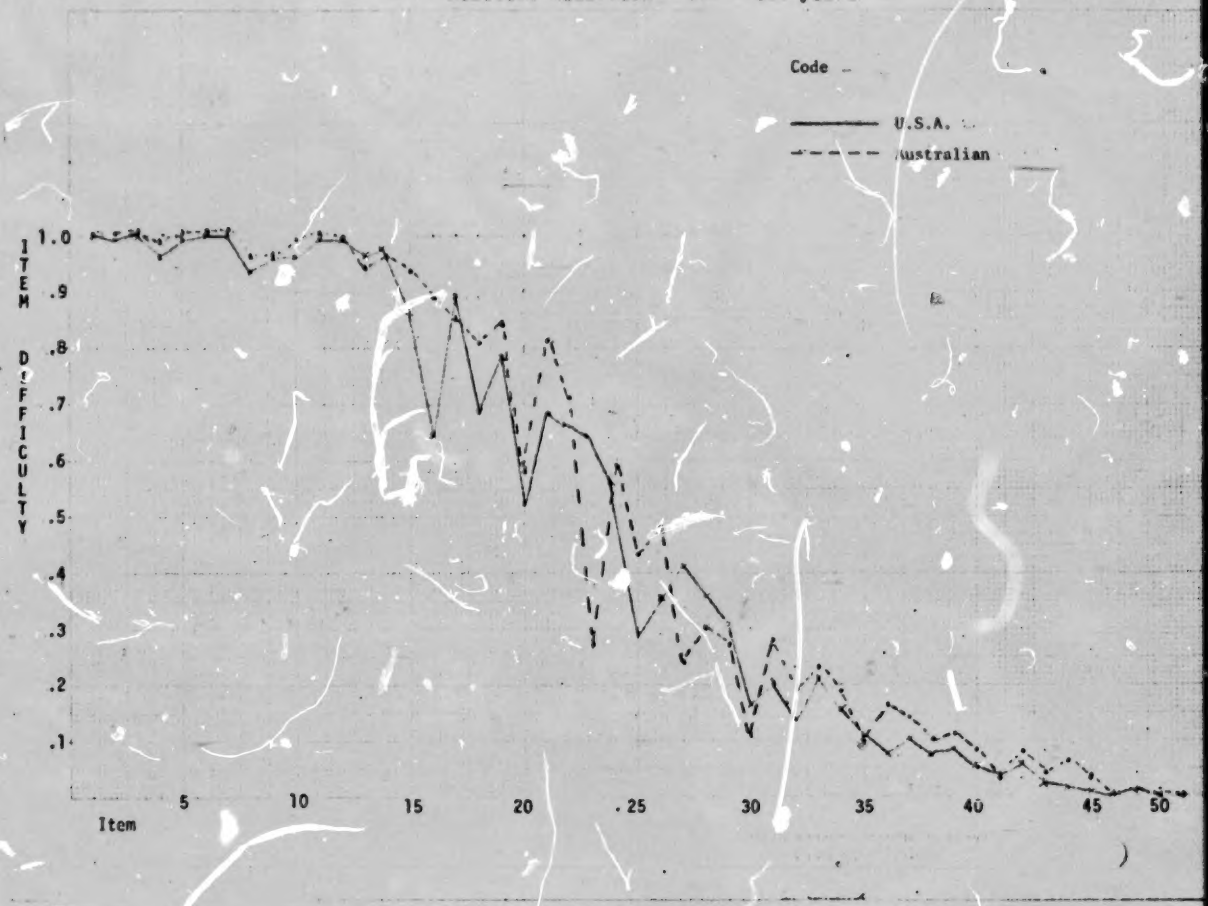

Figure 2
AUDITORY RECEPTION - 6.7 - 7.1 years

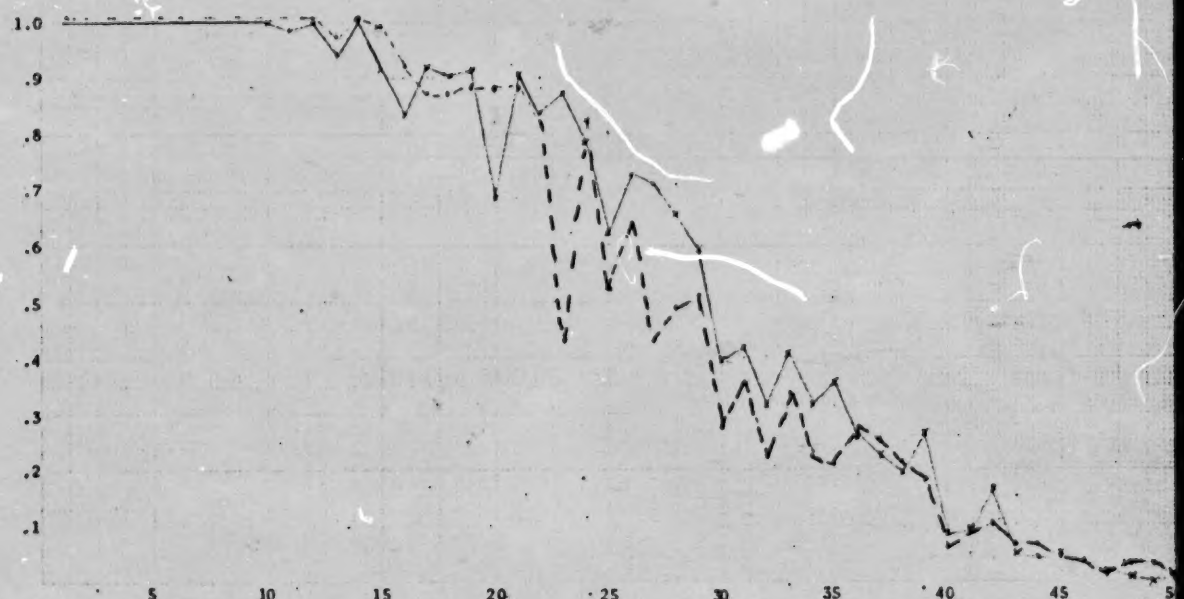

Figure 3
VISUAL RECEPTION - 5.7 - 6.1 years

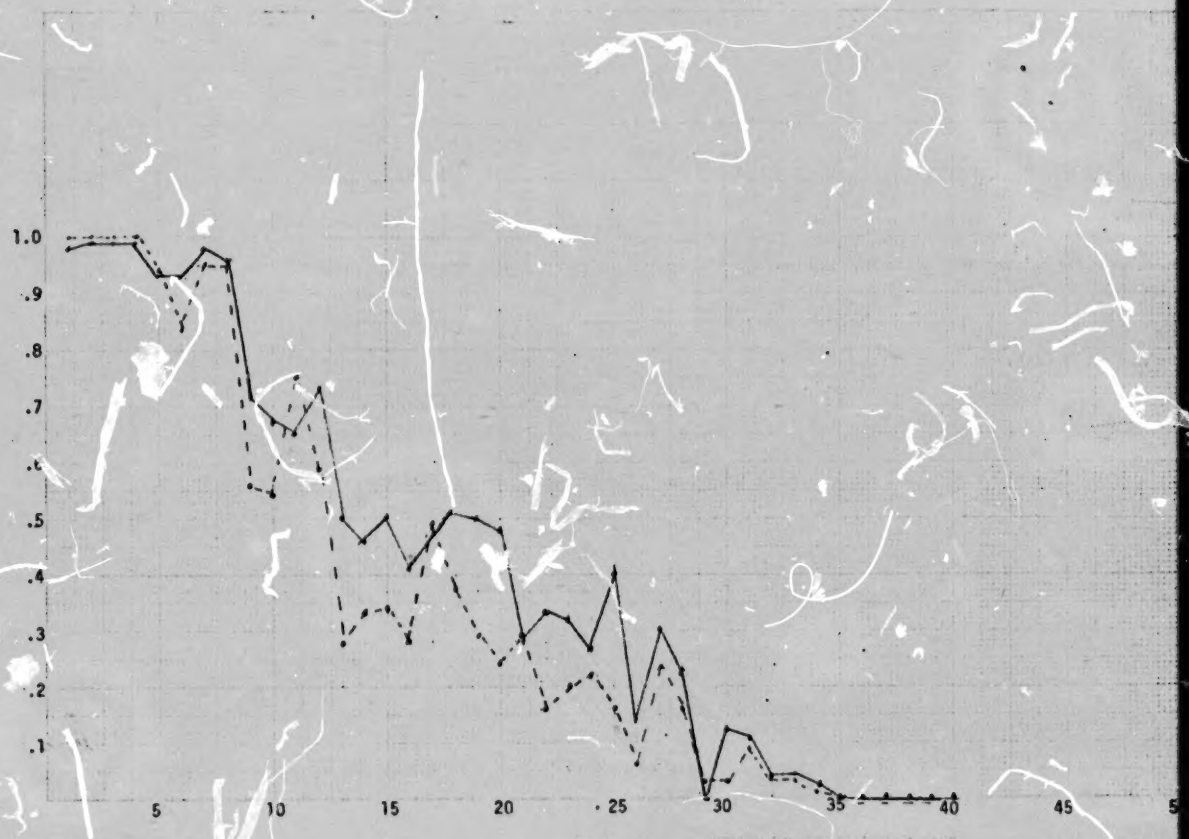

Figure 4
VISUAL RECEPTION - 6.7 - 7.1 years

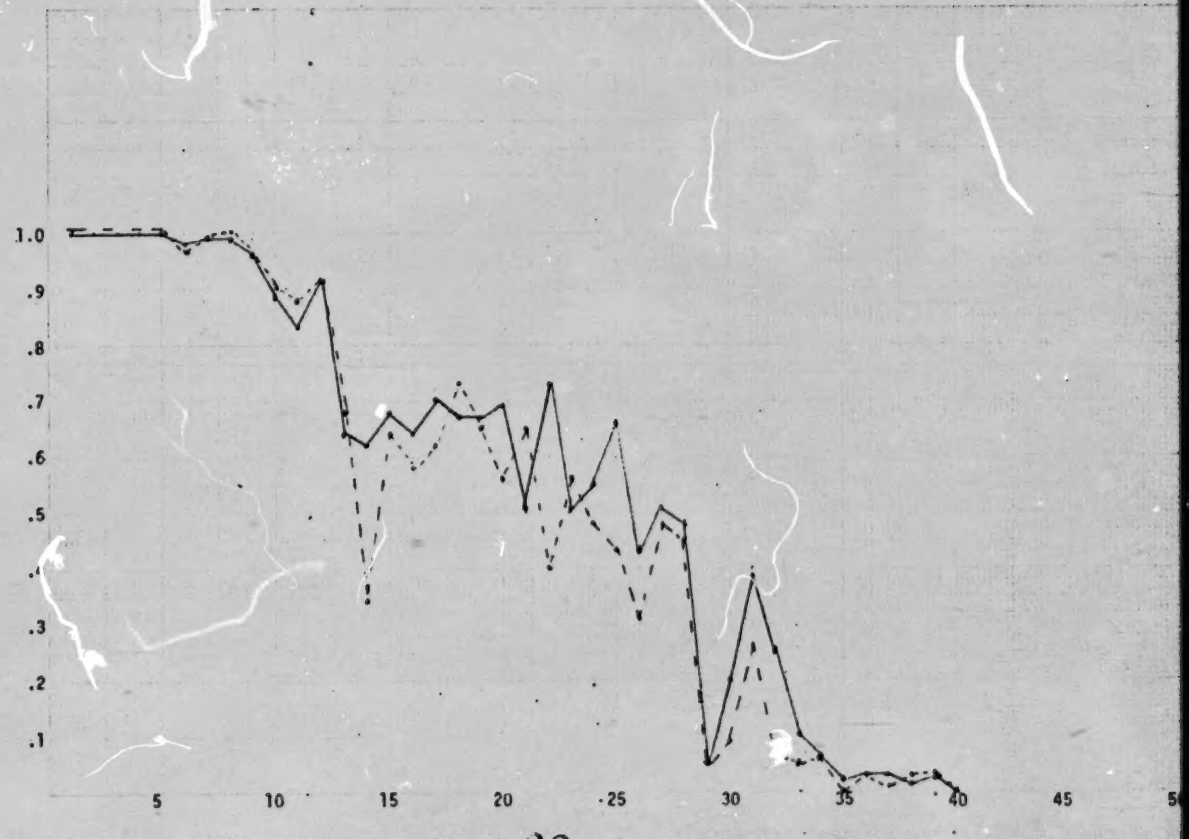

Figure 5
AUDITORY ASSOCIATION - 5.7 - 6.1 years

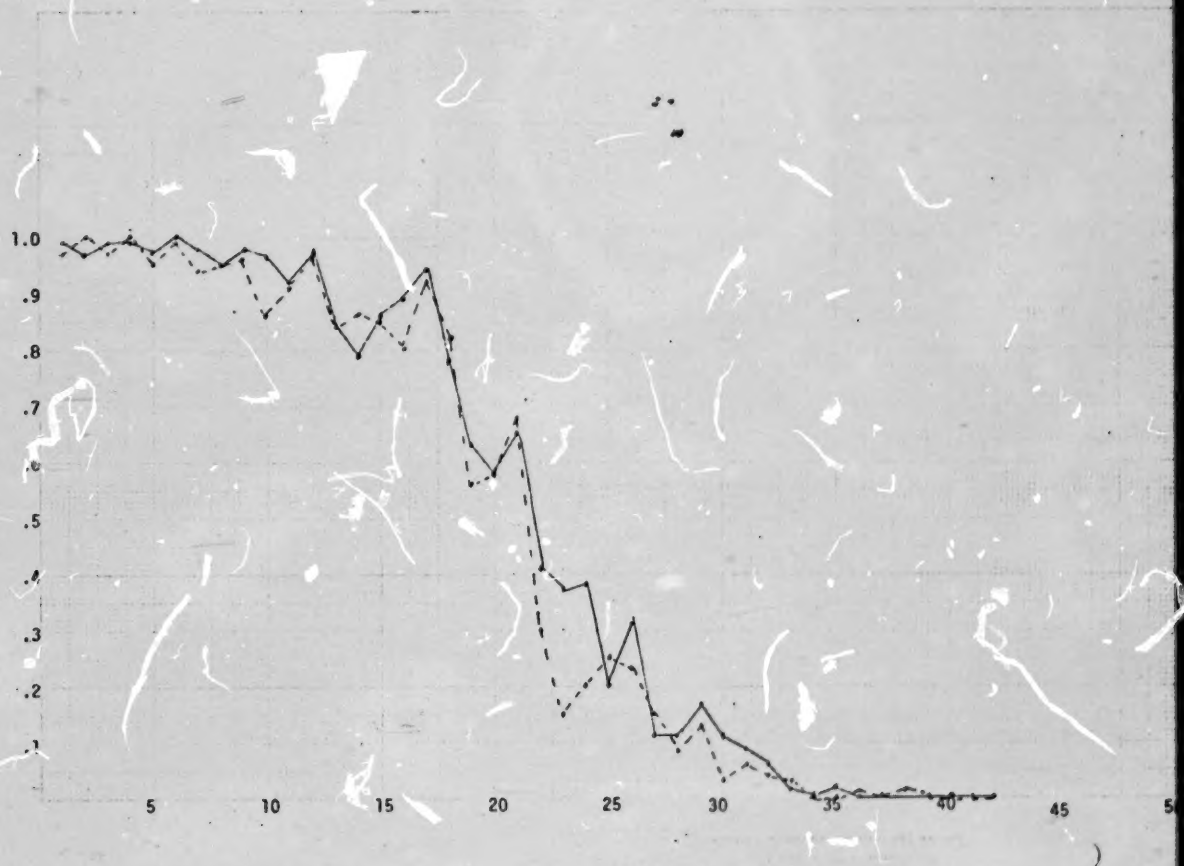

Figure 6
AUDITORY ASSOCIATION - 6.7 - 7.1 years

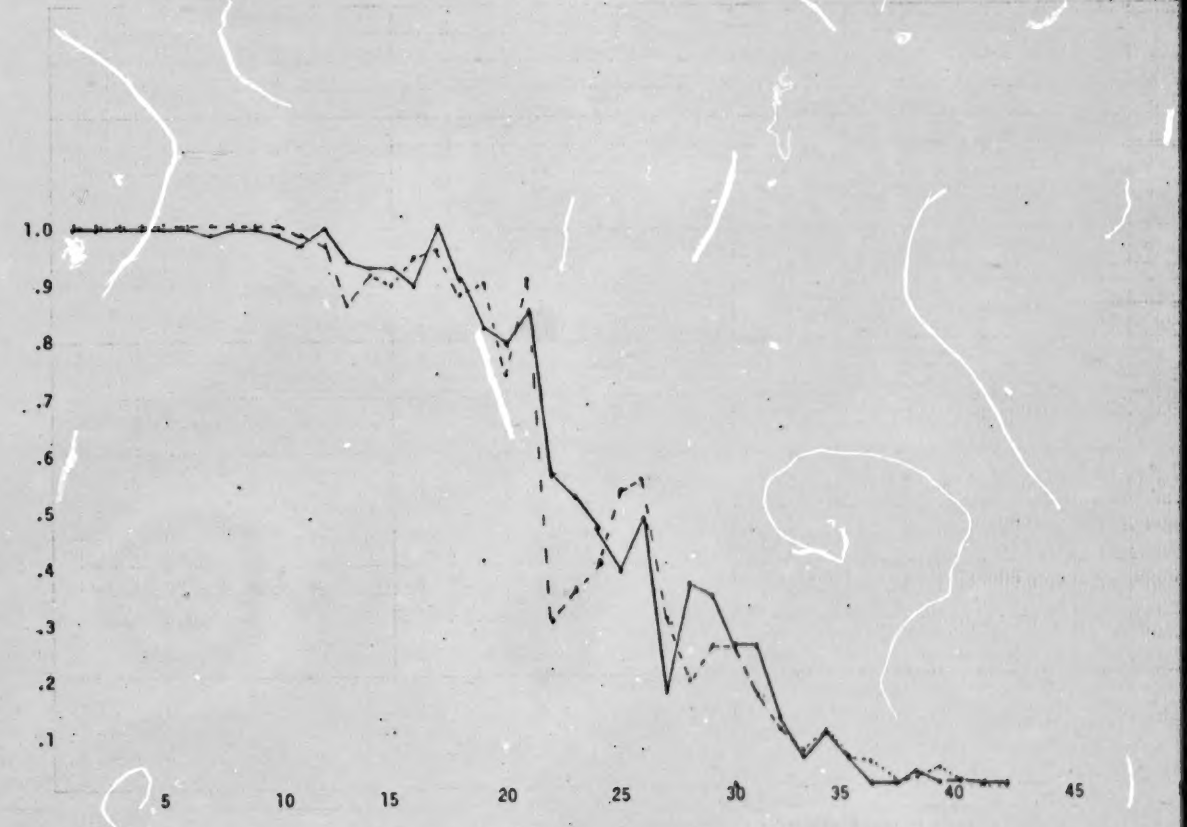

Figure 7
VISUAL ASSOCIATION - 5.7 - 6.7 years

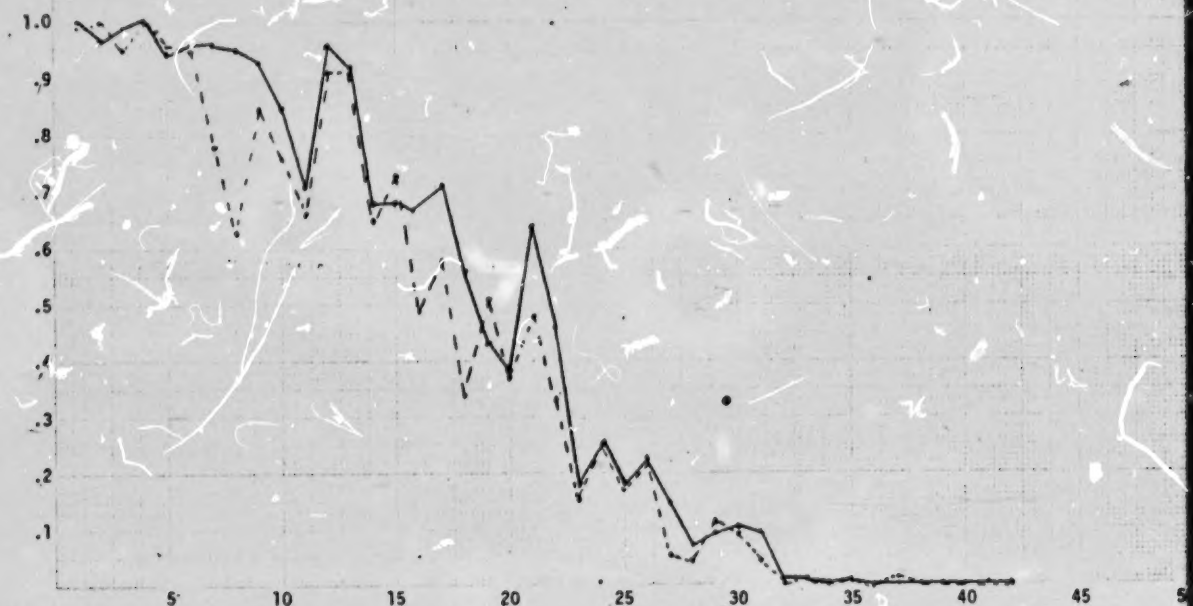

Figure 8
VISUAL ASSOCIATION - 6.7 - 7.1 years

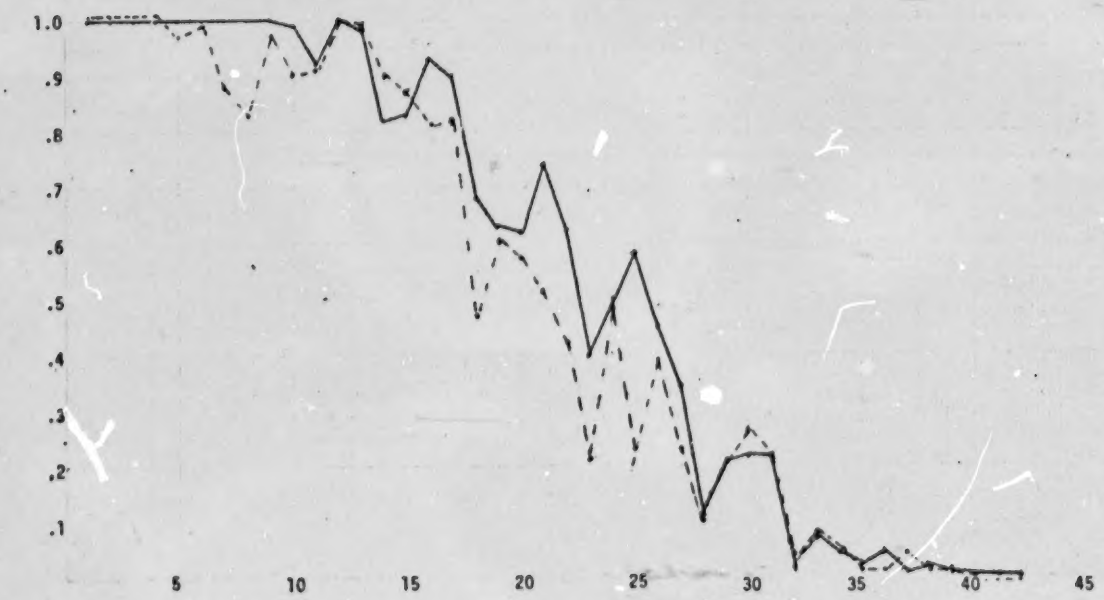

Figure 9
GRAMMATIC CLOSURE - 5.7 - 6.1 years

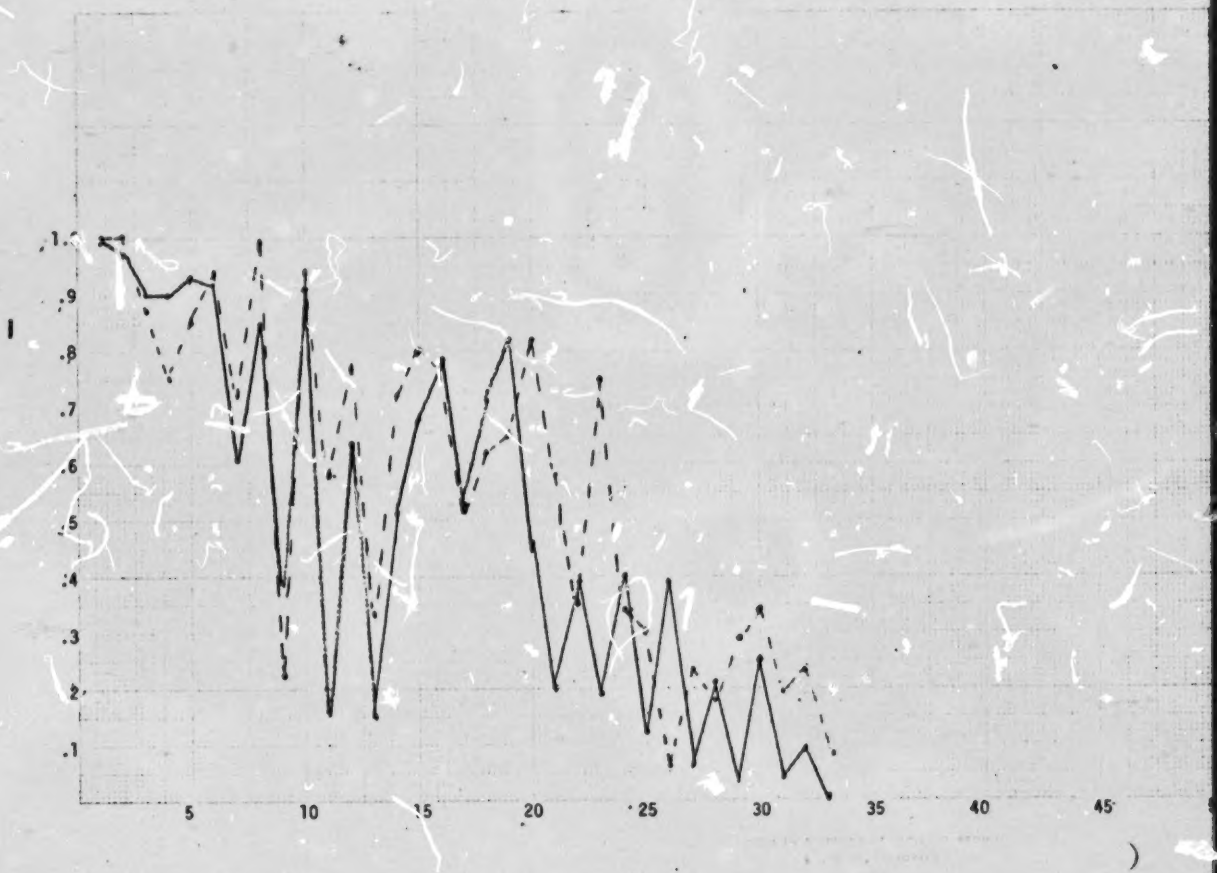

Figure 10
GRAMM[illegible]C CLOSURE - 6.7 - 7.1 years

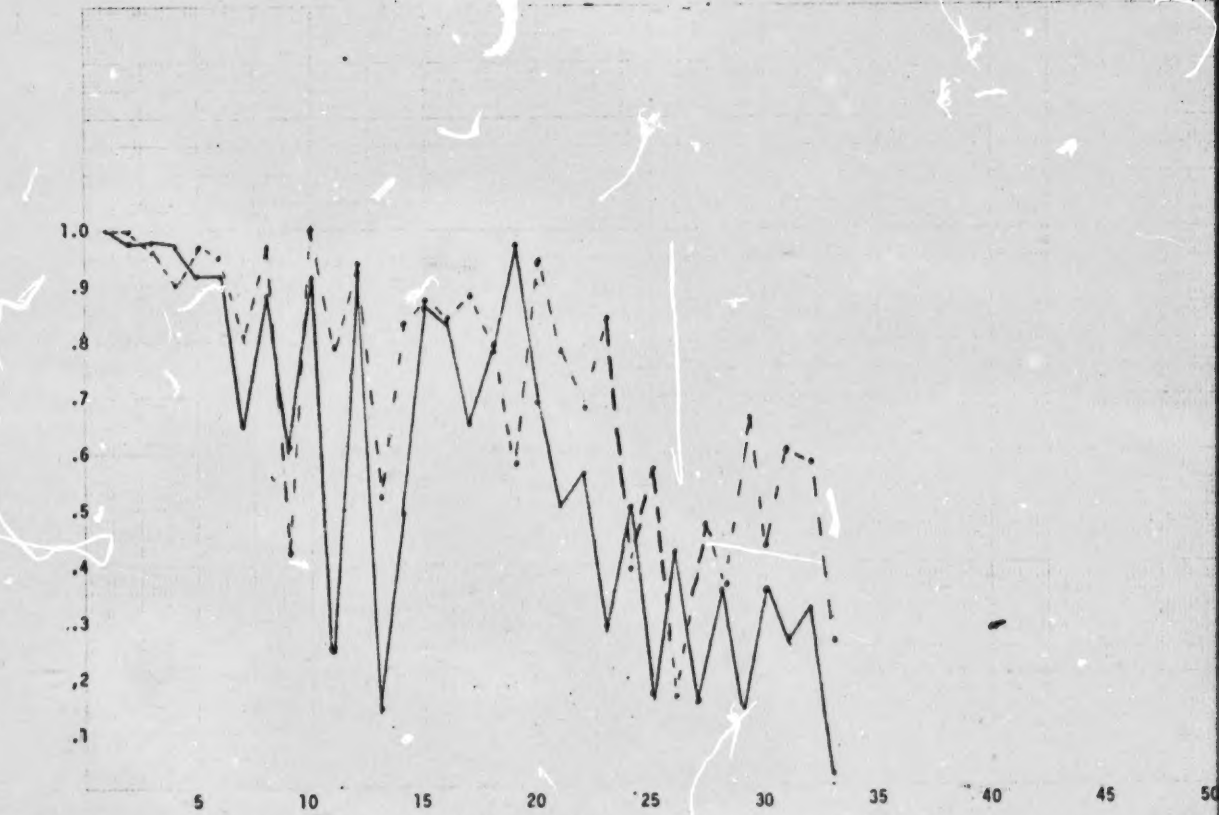

Figure 11

AUDITORY CLOSURE - 5.7 - 6.1 years

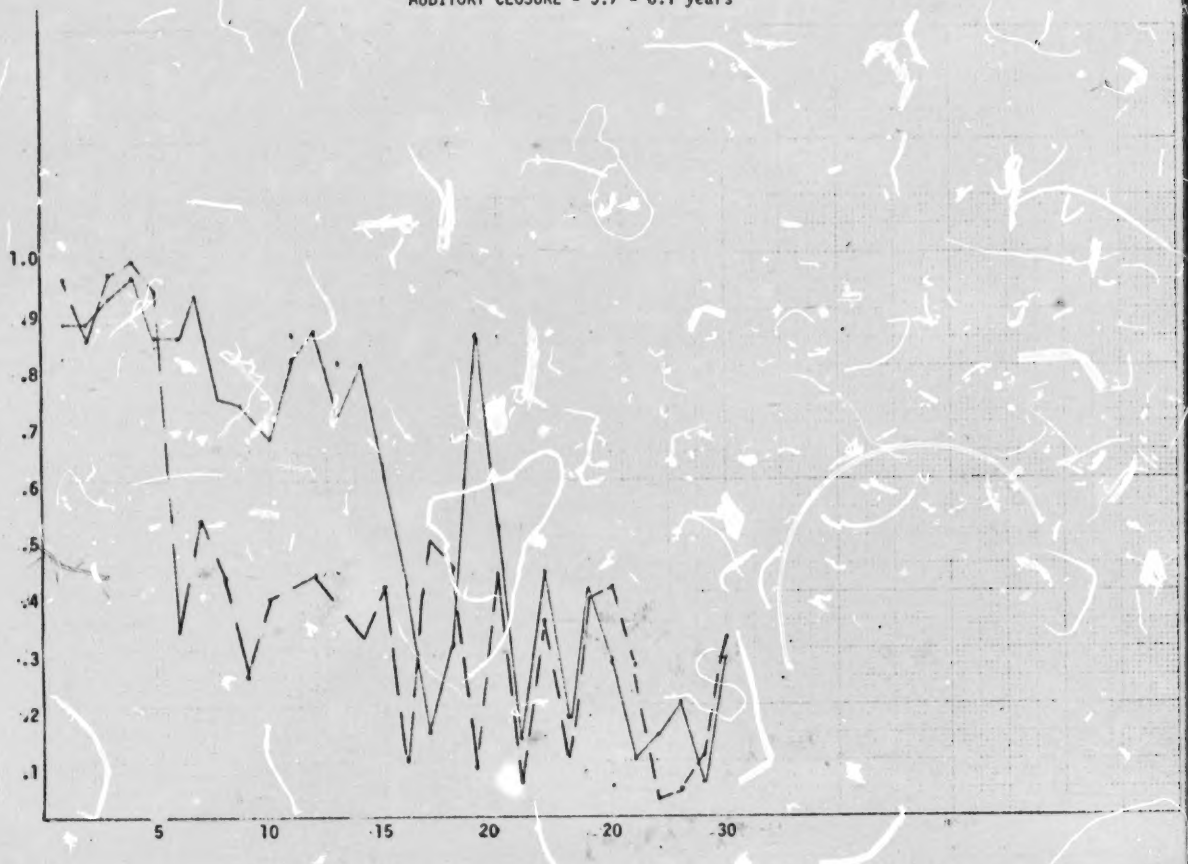

Figure 1

AUDITORY CLOSURE - 7.1 - 7.1 years

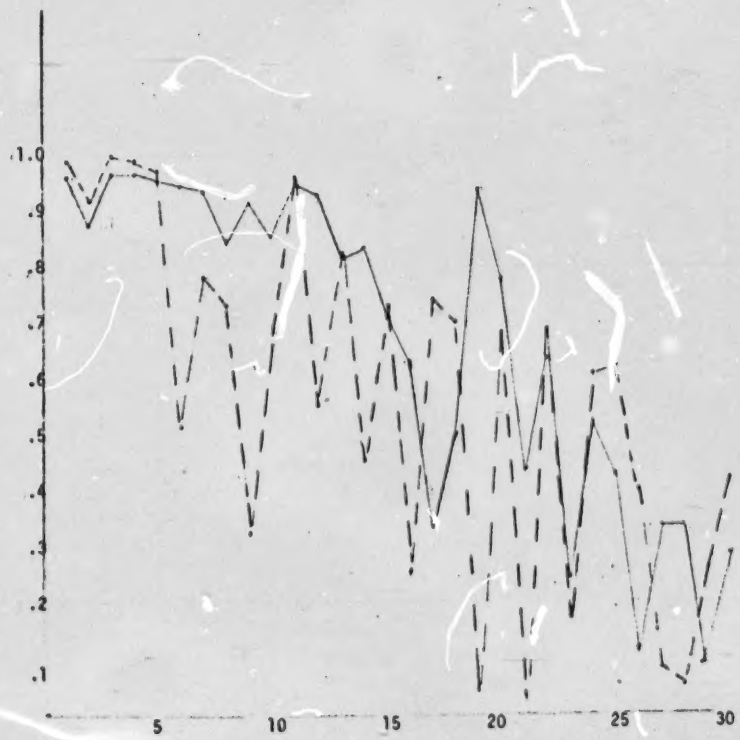

Figure 13
SOUND BLENDING - 5.7 - 6.1 years

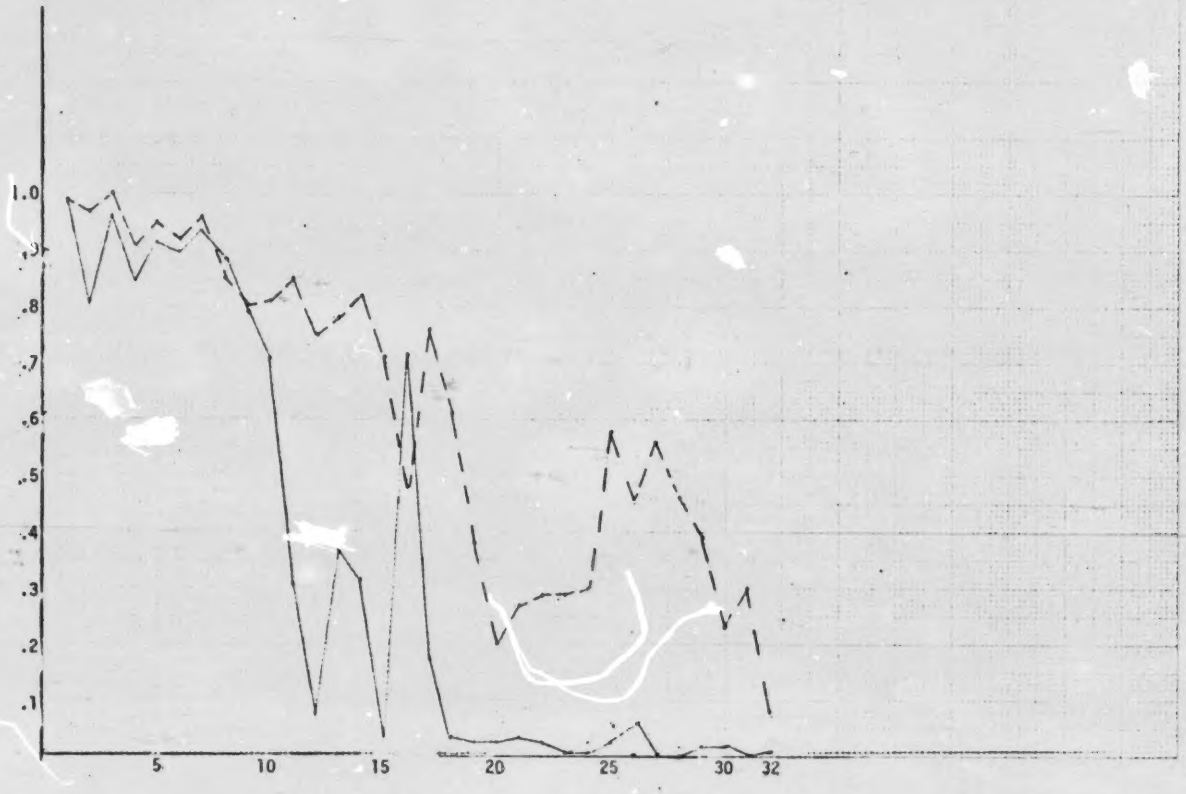

Figure 14
SOUND BLENDING - 6.7 - 7.1 years